Consistency of Bayesian nonparametric inference for discretely observed jump diffusions

Jere Koskela
koskela@math.tu-berlin.de
Institut für Mathematik
Technische Universität Berlin
Straße des 17. Juni 136, 10623 Berlin
Germany

Dario Spanò
d.spano@warwick.ac.uk
Department of Statistics
University of Warwick
Coventry CV4 7AL
UK

Paul A. Jenkins
p.jenkins@warwick.ac.uk
Departments of Statistics and Computer Science
University of Warwick
Coventry CV4 7AL
UK

January 17, 2017

Abstract

We introduce verifiable criteria for weak posterior consistency of identifiable Bayesian nonparametric inference for jump diffusions with unit diffusion coefficient and uniformly Lipschitz drift and jump coefficients in arbitrary dimension. The criteria are expressed in terms of coefficients of the SDEs describing the process, and do not depend on intractable quantities such as transition densities. We also show that products of discrete net and Dirichlet mixture model priors satisfy our conditions, again under an identifiability assumption. This generalises known results by incorporating jumps into previous work on unit diffusions with uniformly Lipschitz drift coefficients.

1 Introduction

Jump diffusions are a broad wide class of stochastic processes encompassing systems undergoing deterministic mean-field dynamics, microscopic diffusion and macroscopic jumps. In this paper we let $\mathbf{X} := (\mathbf{X}_t)_{t \geq 0}$ denote a unit jump diffusion, which can be described as a solution to a stochastic differential equation of the form

$$d\mathbf{X}_t = b(\mathbf{X}_t)dt + d\mathbf{W}_t + c(\mathbf{X}_{t-}, d\mathbf{Z}_t) \tag{1}$$

on a domain $\Omega \subseteq \mathbb{R}^d$ given an initial condition $\mathbf{X}_0 = \mathbf{x}_0$, coefficients $b : \Omega \mapsto \mathbb{R}^d$ and $c : \Omega \times \mathbb{R}_0^d \mapsto \mathbb{R}_0^d$, a d-dimensional Brownian motion $(\mathbf{W}_t)_{t \geq 0}$ and a pure jump Lévy process $(\mathbf{Z}_t)_{t \geq 0}$ on $\mathbb{R}_0^d := \mathbb{R}^d \setminus \{\mathbf{0}\}$ with Lévy measure $M(d\mathbf{z})$ satisfying

$$\int_{\mathbb{R}_0^d} (\|\mathbf{z}\|_2^2 \wedge 1) M(d\mathbf{z}) < \infty$$

The notation $\| \cdot \|_{p,\rho}$ denotes the $L^p(\rho)$-norm, where the Lebesgue measure is meant whenever the measure ρ is omitted.

Jump diffusions are used as models across a broad spectrum of applications, such as economics and finance [Merton, 1976, Aase and Guttorp, 1987, Bardhan and Chao, 1993, Chen and Filipović, 2005, Filipović et al., 2007], biology [Kallianpur, 1992, Kallianpur and Xiong, 1994, Bertoin and Le Gall, 2003, Birkner et al., 2009] and engineering [Au et al., 1982, Bodo et al., 1987]. They also contain many important families of stochastic processes as special cases, including diffusions and Lévy processes.

Remark 1. In the exposition above, the processes $\mathbf{X}$, $\mathbf{W}$ and $\mathbf{Z}$ all share a common dimension. This restriction is not necessary for any of the results in the paper, and has been introduced purely for readability of notation.

Under regularity conditions summarised in the next section, jump diffusions are recurrent, ergodic Feller-Markov processes with transition densities $p_t(\mathbf{x}, \mathbf{y})d\mathbf{y}$ and a unique stationary density $\pi(\mathbf{x})d\mathbf{x}$ with respect to the d-dimensional Lebesgue measure. Under such conditions the procedure of Bayesian inference can be applied to infer the coefficients of the jump diffusion based on observations taken at discrete times. In this paper we focus on joint inference of the drift function b and the family of Lévy measures $\nu(\mathbf{x}, d\mathbf{z}) := M(c^*(\mathbf{x}, d\mathbf{z}))$, where $c^*(\mathbf{x}, \cdot)$ denotes the pull-back of $c(\mathbf{x}, \cdot)$:

$$c^*(\mathbf{x}, d\mathbf{z}) := \{\mathbf{y} \in \Omega : c(\mathbf{x}, \mathbf{y}) \in d\mathbf{z}\}.$$

We abuse terminology and refer to the collection of measures $\nu(\mathbf{x}, \cdot)$ as a Lévy measure for the remainder of the paper. Inference of the Lévy measure will refer to inference of ν, assuming that neither c nor M is known.

More precisely, let Θ denote a set of pairs (b, ν), and let Π denote a prior distribution on $(\Theta, \mathcal{B}(\Theta))$, where $\mathcal{B}(\Theta)$ is the Borel σ-algebra. Let $\mathbf{x}_{0:n} = (\mathbf{x}_0, \mathbf{x}_\delta, \dots, \mathbf{x}_{\delta n})$ denote a time series of observations sampled from a stationary jump diffusion $\mathbf{X}$ at fixed separation δ. The object of interest is the posterior distribution, which can be expressed as

$$\Pi(A|\mathbf{x}_{0:n}) := \frac{\int_A \pi^{b,\nu}(\mathbf{x}_0) \prod_{i=1}^n p_\delta^{b,\nu}(\mathbf{x}_{i-1}, \mathbf{x}_i)\Pi(db, d\nu)}{\int_\Theta \pi^{b,\nu}(\mathbf{x}_0) \prod_{i=1}^n p_\delta^{b,\nu}(\mathbf{x}_{i-1}, \mathbf{x}_i)\Pi(db, d\nu)}$$

for measurable sets $A \in \mathcal{B}(\Theta)$. In the Bayesian setting, the posterior encodes all the available information for inferential purposes. The restriction to unit diffusion coefficients implicit in (1) is a strong assumption in dimension $d > 1$, though some models which fail to satisfy it outright can still be treated via the Lamperti transform [Aït-Sahalia, 2008]. We will outline this procedure briefly in Section 2.

A typical approach to practical Bayesian inference is to choose Θ comprised of parametric families of drift functions and Lévy measures, and fit these parameters to data. However, the natural parameter spaces for jump diffusions are spaces of functions and measures, which are infinite dimensional and cannot be represented in terms of finitely many parameters without significant loss of modelling freedom. Nonparametric Bayesian inference can be thought of as inference of infinitely many parameters, and retains much of the modelling freedom inherent in the class of jump diffusions.

A natural and central question is whether the Bayes procedure is *consistent*, that is, whether the posterior concentrates on a neighbourhood of the parameter space which specifies the "true" dynamics generating the data as the number of observations grows. If $(b_0, \nu_0) \in \Theta$ denotes the data generating drift and Lévy measure, consistency can be expressed as $\Pi(U_{b_0,\nu_0}^c|\mathbf{x}_{0:n}) \to 0$ as $n \to \infty$, where U_{b_0,ν_0} is an open neighbourhood of (b_0, ν_0).

Whether or not Bayesian posterior consistency holds in the nonparametric setting is an intricate question, and depends on subtle ways on the prior Π and the topology endowed on Θ [Diaconis and Freedman, 1986]. A further difficulty in the present context is the fact that stationary and transition densities of jump diffusions are intractable in practically all cases of

interest, so that usual conditions for posterior consistency are difficult to verify. These difficulties were recently overcome for discretely observed, one-dimensional unit diffusions under restrictive conditions on the drift function [van der Meulen and van Zanten, 2013], and a multidimensional generalisation was presented in [Gugushvili and Spreij, 2014]. Both results rely on martingale arguments developed by Ghosal and Tang for Markov processes with tractable transition probabilities [Ghosal and Tang, 2006, Tang and Ghosal, 2007]. A Bayesian analysis of continuously observed one dimensional diffusions has also been conducted under various setups [van der Meulen et al., 2006, Panzar and van Zanten, 2009, Pokern et al., 2013], and a review of Bayesian methods for one dimensional diffusions is provided by [van Zanten, 2013]. Similar developments have also been made for frequentist drift estimation from discrete observations, both for one dimensional unit diffusions [Jacod, 2000, Gobet et al., 2004, Comte et al., 2007] and their multi-dimensional generalisations [Dalalyan and Reiß, 2007, Schmisser, 2013].

The main result of this paper is consistency of Bayesian nonparametric joint inference of drift functions and Lévy measures in arbitrary dimension under verifiable conditions on the prior, given an identifiability assumption which seems difficult to verify in general. This generalises the result of [Gugushvili and Spreij, 2014] by incorporating discontinuous processes with jumps. We also show that products of discrete net and Dirichlet process mixture distributions provide a class of priors for which our conditions hold. The key results enabling this generalisations are a generalised Girsanov-type change of measure theorem for jump diffusions [Cheridito et al., 2005] and a coupling method for establishing regularity of semigroups [Wang, 2010].

The rest of the paper is organised as follows. In Section 2 we introduce the jump diffusion processes in finite dimensional domains and necessary regularity conditions. In Section 3 we define the inference problem under study, and state and prove the corresponding consistency result. In Section 4 we introduce the discrete net prior, and show that it satisfies our consistency conditions. Section 5 concludes with a discussion.

2 Jump diffusions

A general time-homogeneous, d-dimensional jump diffusion $\mathbf{Y} := (\mathbf{Y}_t)_{t\geq 0}$ is the solution of a stochastic differential equation of the form

$$d\mathbf{Y}_t = b(\mathbf{Y}_t)dt + \sigma(\mathbf{Y}_t)d\mathbf{W}_t + c(\mathbf{Y}_{t-}, d\mathbf{Z}_t),$$

where $\sigma : \Omega \mapsto \mathbb{R}^{d\times d}$ and the other coefficients are as in (1). The implicit assumption in (1) of $\sigma \equiv 1$ is restrictive in dimensions $d > 1$. Processes which do not have unit diffusion coefficient can be dealt with provided they lie in the domain of the Lamperti transform [Aït-Sahalia, 2008], i.e. if there exists a mapping $q : \mathbf{Y} \mapsto \mathbf{X}$ such that $\mathbf{X}$ is of the form (1). Such transforms exist for any non-degenerate process in one dimension, but only rarely in higher dimensions. Sufficient conditions for the Lamperti transform to be well defined are non-singularity of σ and the following symmetry condition [Yu, 2007, Aït-Sahalia, 2008]:

$$\frac{\partial(\sigma^{-1})_{ij}(\mathbf{x})}{\partial x_k} = \frac{\partial(\sigma^{-1})_{ik}(\mathbf{x})}{\partial x_j} \text{ for all } i, j, k \in \{1, \dots, d\}. \tag{2}$$

We note also that the Lamperti transform cannot be constructed from discrete data, so that in any case σ must be known a priori. While restrictive, this assumption cannot be relaxed without fundamental changes to the method of proof of consistency and already arises in the simpler case of diffusions without jumps [van der Meulen and van Zanten, 2013, Gugushvili and Spreij, 2014].

The following proposition summarises the necessary regularity assumptions for existence and uniqueness of Feller-Markov jump diffusions with transition densities and a unique stationary density:

Proposition 1. *Assume that $c(\cdot, 0) \equiv 0$, and that there exist constants $C_1, C_2, C_3, C_4 > 0$ such that*

$$\|b(\mathbf{x}) - b(\mathbf{y})\|_2^2 + \int_{\mathbb{R}_0^d} \|c(\mathbf{x}, \mathbf{z}) - c(\mathbf{y}, \mathbf{z})\|_2^2 M(d\mathbf{z}) \leq C_1 \|\mathbf{x} - \mathbf{y}\|_2^2 \tag{3}$$

$$\|c(\mathbf{x}, \mathbf{z}) - c(\mathbf{x}, \boldsymbol{\xi})\|_2^2 \leq C_2 \|\mathbf{z} - \boldsymbol{\xi}\|_2^2 \tag{4}$$

$$\text{For every } \mathbf{x} \in \Omega : \|\mathbf{x}\|_2 > C_3 \text{ the following holds: } \mathbf{x} \cdot b(\mathbf{x}) \leq -C_4 \|\mathbf{x}\|_2^2 \tag{5}$$

$$\int_{\mathbb{R}_0^d : \|\mathbf{z}\|_2 > 1} \|\mathbf{z}\|_2^2 M(dz) < \infty. \tag{6}$$

Then (1) has a unique, ergodic weak solution $\mathbf{X}$ with the Feller and Markov properties. Furthermore, $\mathbf{X}$ has a unique stationary density $\pi^{b,\nu}(\mathbf{x})d\mathbf{x}$ with a finite second moment, and the associated semigroup $P_t^{b,\nu}$ has transition densities $p_t^{b,\nu}(\mathbf{x}, \mathbf{y})d\mathbf{y}$.

Proof. Existence and uniqueness of $\mathbf{X}$ are obtained from (3), as well as the linear growth bounds implied by Lipschitz continuity, by Theorem 6.2.9 of [Applebaum, 2004]. Theorem 6.4.6 of [Applebaum, 2004] gives the Markov property under the same conditions. Finally, the corollary in Appendix 1 of [Kolokoltsov, 2004] yields the Feller property. In turn, the Feller property and the fact that $\log(1 + \|\boldsymbol{\xi}\|_2)^{-1} \|\boldsymbol{\xi}\|_2^2 \to \infty$ as $\|\boldsymbol{\xi}\|_2 \to \infty$ mean that the hypotheses of Theorem 1.1 of [Schilling and Wang, 2013] are fulfilled, so that $\mathbf{X}$ has bounded transition densities with respect to the Lebesgue measure.

Existence and uniqueness of $\pi^{b,\nu}$, as well as ergodicity of $\mathbf{X}$ will follow from Theorem 2.1 of [Masuda, 2007], the hypotheses of which will now be verified. Along with $c(\cdot, 0) \equiv 0$, conditions (3) and (4) above imply Assumption 1 of [Masuda, 2007]. Now, for every $u \in (0, 1)$ let

$$b^u(\mathbf{x}) := b(\mathbf{x}) - \int_{u < \|\mathbf{z}\|_1 \leq 1} c(\mathbf{x}, \mathbf{z}) M(d\mathbf{z}).$$

Assumption 2(a)' of [Masuda, 2009] requires $\mathbf{X}$ to admit bounded transition densities, and the diffusion which solves

$$d\mathbf{X}_t^u = b^u(\mathbf{X}_t^u)dt + \sigma(\mathbf{X}_t^u)d\mathbf{W}_t$$

to be irreducible for each $u > 0$. Boundedness of the transition density of $\mathbf{X}$ was established above, and irreducibility of $\mathbf{X}^u$ holds because $\sigma \equiv 1$ by Theorem 2.3 of [Stramer and Tweedie, 1997].

Next we verify Assumptions 3 and 3* of [Masuda, 2007] by checking the conditions of Lemma 2.4' of [Masuda, 2009]. The diffusion coefficient is constant, and hence $o(\|x\|_2^{1-q/2})$ for any $q \in (0, 2)$. Condition (6) is the corresponding hypothesis of [Masuda, 2009], and both $\|\mathbf{x}\|_2^{q-2} \mathbf{x} \cdot b(\mathbf{x}) \to -\infty$ and $\|\mathbf{x}\|_2^{-2} \mathbf{x} \cdot b(\mathbf{x}) \leq -C_4$ follow from (5). Hence, Assumptions 3 and 3* of [Masuda, 2007] hold. This yields ergodicity (and mixing) by Theorem 2.1 of [Masuda, 2007], and second moments of the stationary distribution (and exponential mixing) by Theorem 2.2 of [Masuda, 2007].

It remains to show the invariant measure has a density. By combining Proposition 5.1.9 and Theorem 5.1.8 of [Fornaro, 2004] it can be seen that invariant measures of irreducible strong Feller processes are equivalent to the associated transition probabilities, which is sufficient in this case. Assumption 1 of [Masuda, 2007] and Assumption 2(a)' of [Masuda, 2009] imply irreducibility of $\mathbf{X}$ (c.f. Claim 1 on page 42 of [Masuda, 2007]). Condition (3) guarantees the strong Feller property by Theorem 2.3 of [Wang, 2010]. Hence the invariant measure has a density with respect to the transition densities, and thus also the Lebesgue measure. This concludes the proof. □

Remark 2. Assumption (3) is central to the proof of our main result. In contrast, assumptions (4), (5) and (6) are only needed to ensure the conclusions of Proposition 1, as well as (10) and

(11). Whenever the prior Π is supported by a set Θ for which versions of Proposition 1, (10) and (11) can be established without (4), (5) and (6), these assumptions can be discarded without affecting our results.

We denote the law of $\mathbf{X}$ with drift function b, Lévy measure ν and initial condition $\mathbf{X}_0 = \mathbf{x}$ by $\mathbb{P}_{\mathbf{x}}^{b,\nu}$ and the corresponding expectation by $\mathbb{E}_{\mathbf{x}}^{b,\nu}$. Dependence on initial conditions is omitted when the stationary process is meant.

3 Consistency for discrete observations

We begin by defining the topology and weak posterior consistency following the set up of [van der Meulen and van Zanten, 2013]. In addition to topological details, posterior consistency is highly sensitive to the support of the prior, which should not exclude the truth. This is guaranteed by insisting that the prior places positive mass on all neighbourhoods of the truth, typically measured in terms of Kullback-Leibler divergence. In our setting such a support condition is provided by (8) below.

We begin by setting out the necessary assumptions on the parameter space Θ.

Definition 1. Let $\Theta = \{(b, \nu) : b : \Omega \mapsto \mathbb{R}^d, \nu : \Omega \times \mathbb{R}_0^d \mapsto \mathbb{R}_+\}$ denote a set of pairs of drift functions $b(\mathbf{x})$ and Lévy measures $\nu(\mathbf{x}, d\mathbf{z}) := M(c^*(\mathbf{x}, d\mathbf{z}))$ with each pair satisfying the hypotheses of Proposition 1. Furthermore, suppose that there exists a constant $C_5 > 0$ such that

$$M(\{\mathbf{z} \in \Omega : \|\mathbf{z}\|_2 > 1\}) < C_5 \tag{7}$$

uniformly in Θ, and that for each $\mathbf{x} \in \Omega$ and any pair of Lévy measures $(\cdot, \nu), (\cdot, \nu') \in \Theta$ the measures $\nu(\mathbf{x}, \cdot) \sim \nu'(\mathbf{x}, \cdot)$ are equivalent with strictly positive, finite Radon-Nikodym density $0 < \frac{d\nu'}{d\nu} < \infty$, and that either

1. $\nu(\mathbf{x}, \cdot)$ is a finite measure or
2. there exists an open set A containing the origin such that $\nu(\mathbf{x}, \cdot)|_A = \nu'(\mathbf{x}, \cdot)|_A$.

Remark 3. In effect, the conditions of Definition 1 mean that the unit diffusion coefficient and the infinite intensity component of the Lévy measure can be thought of as known confounders of the joint inference problem for the drift function and the compound Poisson component of the Lévy measure driving macroscopic jumps. In particular, one of conditions 1. or 2. is needed to ensure finiteness of the integrals (10) and (11).

The following assumption ensures that the drift function and Lévy measure can be uniquely identified from discrete data. While it is crucial for our main result — posterior consistency as we will define it below does not even make sense for a non-identifiable inference problem — we have not been able to produce a tractable identifiability condition beyond that given in [Gugushvili and Spreij, 2014] for gradient-type diffusions. See also page 398 of [Bladt and Sørensen, 2005] for a discussion on the challenge posed by identifiability in the simpler setting of discretely observed Markov processes with finite state spaces.

Assumption 1. For Π-almost any pair $(b, \nu) \neq (b', \nu') \in \Theta$ there exists $\mathbf{x} \in \Omega$ and $f \in D(G^{b,\nu})$ such that $P_\delta^{b,\nu} f(\mathbf{x}) \neq P_\delta^{b',\nu'} f(\mathbf{x})$. In particular, identifying $P_\delta^{b,\nu}$ is equivalent to identifying (b, ν). We emphasize that both $\mathbf{x}$ and f may depend on (b, ν) and (b', ν').

The topology under consideration is defined as in [van der Meulen and van Zanten, 2013, Gugushvili and Spreij, 2014] by specifying a subbase determined by the semigroups $P_t^{b,\nu}$. For details about the notion of a subbase, and other topological concepts, see e.g. [Dudley, 2002].

Definition 2. Fix a sampling interval $\delta > 0$ and a finite measure $\rho \in \mathcal{M}_f(\Omega)$ with positive mass in all non-empty, open sets. For any $(b, \nu) \in \Theta$, $\varepsilon > 0$ and $f \in C_b(\Omega)$ define the set

$$U_{f,\varepsilon}^{b,\nu} := \{(b', \nu') \in \Theta : \| P_\delta^{b',\nu'} f - P_\delta^{b,\nu} f \|_{1,\rho} < \varepsilon \}.$$

A weak topology on Θ is generated by requiring that the family $\{U_{f,\varepsilon}^{b,\nu} : f \in C_b(\Omega), \varepsilon > 0, (b, \nu) \in \Theta\}$ is a subbase of the topology.

The following lemma is a direct analogue of Lemma 3.2 of [van der Meulen and van Zanten, 2013]:

Lemma 1. *The topology generated by a subbase of sets of the form $U_{f,\varepsilon}^{b,\nu}$ is Hausdorff.*

Proof. Consider $(b, \nu) \neq (b', \nu') \in \Theta$. By Assumption 1 there exists $f \in C(\Omega)$ and $\mathbf{x} \in \Omega$ such that $P_\delta^{b,\nu} f(\mathbf{x}) \neq P_\delta^{b',\nu'} f(\mathbf{x})$, and hence by continuity a nonempty open set $J \subset \Omega$ where $P_\delta^{b,\nu} f$ and $P_\delta^{b',\nu'} f$ differ. Hence $\| P_\delta^{b,\nu} f - P_\delta^{b',\nu'} f \|_{1,\rho} > \varepsilon$ for some $\varepsilon > 0$ so that the neighbourhoods $U_{f,\varepsilon/2}^{b,\nu}$ and $U_{f,\varepsilon/2}^{b',\nu'}$ are disjoint. $\square$

We are now in a position to formally define posterior consistency, and state the main result of the paper.

Definition 3. Let $\mathbf{x}_{0:n} := (\mathbf{x}_0, \ldots, \mathbf{x}_n)$ denote $n + 1$ samples observed at sampling times $0, \delta, \ldots, \delta n$ from $\mathbf{X}$ at stationarity, i.e. with initial distribution $\mathbf{X}_0 \sim \pi^{b_0,\nu_0}$. Weak posterior consistency holds if $\Pi(U_{b_0,\nu_0}^c | \mathbf{x}_{0:n}) \to 0$ with $\mathbb{P}^{b_0,\nu_0}$-probability 1 as $n \to \infty$, where U_{b_0,ν_0} is any open neighbourhood of $(b_0, \nu_0) \in \Theta$.

Theorem 1. *Let $\mathbf{x}_{0:n}$ be as in Definition 3, and suppose that the prior Π is supported on a set Θ which satisfies Assumption 1 and the conditions of Definition 1, with the constants C_1 and C_5 in (3) and (6) holding uniformly in Θ. If*

$$\Pi\Bigg((b, \nu) \in \Theta : \frac{1}{2} \left(\| b_0 - b \|_{2,\pi^{b_0,\nu_0}} + \left\| \int_{\mathbb{R}_0^d} \left[\frac{d\nu_0}{d\nu}(\cdot, \mathbf{z}) - 1 \right] \mathbb{1}_{(0,1]}(\|\mathbf{z}\|_2) \mathbf{z} \nu(\cdot, d\mathbf{z}) \right\|_{2,\pi^{b_0,\nu_0}} \right)^2 + \left\| \int_{\mathbb{R}_0^d} \left[\log\left(\frac{d\nu_0}{d\nu}(\cdot, \mathbf{z}) \right) - \frac{d\nu_0}{d\nu}(\cdot, \mathbf{z}) + 1 \right] \nu_0(\cdot, d\mathbf{z}) \right\|_{1,\pi^{b_0,\nu_0}} < \varepsilon \Bigg) > 0 \tag{8}$$

for any $\varepsilon > 0$ and any $(b_0, \nu_0) \in \Theta$, then weak posterior consistency holds for Π on Θ.

Proof. We prove Theorem 1 by generalising the proof of Theorem 3.5 of [van der Meulen and van Zanten, 2013]. For $(b, \nu) \in \Theta$ let $\mathrm{KL}(b_0, \nu_0; b, \nu)$ denote the Kullback-Leibler divergence between $p_\delta^{b_0,\nu_0}$ and $p_\delta^{b,\nu}$:

$$\mathrm{KL}(b_0, \nu_0; b, \nu) := \int_\Omega \int_\Omega \log\left(\frac{p_\delta^{b_0,\nu_0}(\mathbf{x}, \mathbf{y})}{p_\delta^{b,\nu}(\mathbf{x}, \mathbf{y})} \right) p_\delta^{b_0,\nu_0}(\mathbf{x}, \mathbf{y}) \pi^{b_0,\nu_0}(\mathbf{x}) d\mathbf{y} d\mathbf{x},$$

and for two probability measures P, P' on the same σ-field let $K(P, P') := \mathbb{E}_P \left[\log\left(\frac{dP}{dP'} \right) \right]$. The law of a random object Z under a probability measure P is denoted by $\mathcal{L}(Z|P)$.

We require the following two properties:

1. $\Pi((b, \nu) \in \Theta : \mathrm{KL}(b_0, \nu_0; b, \nu) < \varepsilon) > 0$ for any $\varepsilon > 0$.
2. Uniform equicontinuity of the functions $\{P_\delta^{b,\nu} f : (b, \nu) \in \Theta\}$ for $f \in C_b(\Omega)$, the set of bounded, continuous functions on Ω.

These two properties will be established in Lemmas 2 and 3 below, which are the necessary generalisations of Lemmas 5.1 and A.1 of [van der Meulen and van Zanten, 2013], respectively.

Lemma 2. *Condition* (8) *implies that* $\Pi((b,\nu) \in \Theta : \mathrm{KL}(b_0, \nu_0; b, \nu) < \varepsilon) > 0$ *for any* $\varepsilon > 0$.

Proof. As in Lemma 5.1 of [van der Meulen and van Zanten, 2013] it will be sufficient to bound $\mathrm{KL}(b_0, \nu_0; b, \nu)$ from above by a constant multiple of

$$\frac{1}{2}\left(\|b_0 - b\|_{2,\pi^{b_0,\nu_0}} + \left\| \int_{\mathbb{R}_0^d} \left[\frac{d\nu_0}{d\nu}(\cdot, \mathbf{z}) - 1 \right] \mathbb{1}_{(0,1]}(\|\mathbf{z}\|_2)\mathbf{z}\nu(\cdot, d\mathbf{z}) \right\|_{2,\pi^{b_0,\nu_0}} \right)^2$$
$$+ \left\| \int_{\mathbb{R}_0^d} \left[\log\left(\frac{d\nu_0}{d\nu}(\cdot, \mathbf{z}) \right) - \frac{d\nu_0}{d\nu}(\cdot, \mathbf{z}) + 1 \right] \nu_0(\cdot, d\mathbf{z}) \right\|_{1,\pi^{b_0,\nu_0}}.$$

A formal calculation yields

$$\begin{aligned}
&\int_\Omega \int_\Omega \log\left(\frac{\pi^{b_0,\nu_0}(\mathbf{x}) p_\delta^{b_0,\nu_0}(\mathbf{x},\mathbf{y})}{\pi^{b,\nu}(\mathbf{x}) p_\delta^{b,\nu}(\mathbf{x},\mathbf{y})} \right) p_\delta^{b_0,\nu_0}(\mathbf{x},\mathbf{y}) \pi^{b_0,\nu_0}(\mathbf{x}) d\mathbf{y} d\mathbf{x} \\
&= K(\pi^{b_0,\nu_0}, \pi^{b,\nu}) + \mathrm{KL}(b_0,\nu_0; b,\nu) = K(\mathcal{L}(\mathbf{X}_0, \mathbf{X}_\delta | \mathbb{P}^{b_0,\nu_0}), \mathcal{L}(\mathbf{X}_0, \mathbf{X}_\delta | \mathbb{P}^{b,\nu})) \\
&\leq K(\mathcal{L}((\mathbf{X}_t)_{t\in[0,\delta]} | \mathbb{P}^{b_0,\nu_0}), \mathcal{L}((\mathbf{X}_t)_{t\in[0,\delta]} | \mathbb{P}^{b,\nu})) \\
&= K(\pi^{b_0,\nu_0}, \pi^{b,\nu}) + \mathbb{E}^{b_0,\nu_0}\left[\log\left(\frac{d\mathbb{P}_{\mathbf{X}_0}^{b_0,\nu_0}}{d\mathbb{P}_{\mathbf{X}_0}^{b,\nu}}((\mathbf{X}_t)_{t\in[0,\delta]}) \right) \right]
\end{aligned} \tag{9}$$

by the conditional version of Jensen's inequality.

The aim is to identify the Radon-Nikodym derivative using Theorem 2.4 of [Cheridito et al., 2005], the hypotheses of which will now be verified. The local boundedness assumptions of [Cheridito et al., 2005] follow from Lipschitz continuity (3). Moreover, let $\{\Omega_n\}_{n=1}^\infty$ denote a sequence of bounded, open subsets of Ω such that $\Omega_1 \subset \Omega_2 \subset \ldots$ and $\cup_{n\geq 1}\Omega_n = \Omega$. Then Lipschitz continuity, and the assumed finiteness of the Radon-Nikodym derivatives in Definition 1 ensure that there exists a sequence of finite constants $\{K_N\}_{n=1}^\infty$ such that

$$\sup_{\mathbf{x}\in\Omega_n} \left\{ \left\| b_0(\mathbf{x}) - b(\mathbf{x}) - \int_{\mathbb{R}_0^d} \left[\frac{d\nu_0}{d\nu}(\mathbf{x}, \mathbf{z}) - 1 \right] \mathbb{1}_{(0,1]}(\|\mathbf{z}\|_2)\mathbf{z}\nu(\mathbf{x}, d\mathbf{z}) \right\|_2 \right\} < K_n \tag{10}$$

$$\sup_{\mathbf{x}\in\Omega_n} \left\{ \int_{\mathbb{R}_0^d} \left[\frac{d\nu_0}{d\nu}(\mathbf{x}, \mathbf{z}) \log\left(\frac{d\nu_0}{d\nu}(\mathbf{x}, \mathbf{z}) \right) - \frac{d\nu_0}{d\nu}(\mathbf{x}, \mathbf{z}) + 1 \right] \nu(\mathbf{x}, d\mathbf{z}) \right\} < K_n \tag{11}$$

for each $(b,\nu) \in \Theta$ and each $n \in \mathbb{N}$. In particular, the conditions in Remark 2.5 of [Cheridito et al., 2005] are satisfied. Hence Theorem 2.4 of [Cheridito et al., 2005] holds, and the Radon-Nikodym derivative on the RHS of (9) can be expressed as $\mathbb{E}^{b_0,\nu_0}[\log(\mathcal{E}(L_\delta))]$, where $\mathcal{E}$ is the Doléans-Dade stochastic exponential and the process $L := (L_t)_{t\in[0,\delta]}$ is given as

$$\begin{aligned}
L_t = &\int_0^t \int_{\mathbb{R}_0^d} \left[\frac{d\nu_0}{d\nu}(\mathbf{X}_{s-}, \mathbf{z}) - 1 \right] (\mathbf{Z}^\nu(\mathbf{X}_{s-}, d\mathbf{z}, ds) - \nu(\mathbf{X}_{s-}, d\mathbf{z})ds) \\
&+ \int_0^t b_0(\mathbf{X}_s) - b(\mathbf{X}_s) - \int_{\mathbb{R}_0^d} \left(\frac{d\nu_0}{d\nu}(\mathbf{X}_{s-}, \mathbf{z}) - 1 \right) \mathbb{1}_{(0,1]}(\|\mathbf{z}\|_2)\mathbf{z}\nu(\mathbf{X}_{s-}, d\mathbf{z}) d\mathbf{X}_s^c,
\end{aligned}$$

where $(\mathbf{X}_s^c)_{s\geq 0}$ is the continuous martingale part of $\mathbf{X}$, i.e. a Brownian motion in this setting, and $\mathbf{Z}^\nu(\mathbf{x}, \cdot, \cdot)$ is a Poisson random measure with intensity $\nu(\mathbf{x}, d\mathbf{z}) \otimes ds$. Note that under $\mathbb{P}^{b_0,\nu_0}$ the process L is a local martingale, L^c is a continuous local martingale with quadratic variation

$$\langle L^c \rangle_t = \int_0^t \left\| b_0(\mathbf{X}_s) - b(\mathbf{X}_s) - \int_{\mathbb{R}_0^d} \left(\frac{d\nu_0}{d\nu}(\mathbf{X}_{s-}, \mathbf{z}) - 1 \right) \mathbb{1}_{(0,1]}(\|\mathbf{z}\|_2)\mathbf{z}\nu(\mathbf{X}_{s-}, d\mathbf{z}) \right\|_2^2 ds$$

and jump discontinuities of L can be written as

$$\Delta L_t = \left[\frac{d\nu_0}{d\nu}(\mathbf{X}_{t-}, \Delta\mathbf{X}_t) - 1\right] \mathbb{1}_{(0,\infty)}(\|\Delta\mathbf{X}_t\|_2),$$

where $\Delta\mathbf{X}_t$ denotes a jump discontinuity of $\mathbf{X}$ at time t. Now, the expected quadratic variation of $\langle L^c\rangle_t$ can be bounded by

$$\mathbb{E}^{b_0,\nu_0}[\langle L^c\rangle_t] \leq \int_0^t \mathbb{E}^{b_0,\nu_0}[\|b_0(\mathbf{0}) + b(\mathbf{0}) + 2C_1\mathbf{X}_s + K\|_2^2]ds$$

for some constant $K > 0$, using (3), the uniform upper and lower bounds on $\frac{d\nu_0}{d\nu}$, and the fact that either ν and ν_0 are equivalent and either both finite, or $\frac{d\nu_0}{d\nu} \equiv 1$ on a neighbourhood of 0 and ν is finite on any open set not containing the origin. The stationary density has a first moment by Proposition 1, so that $\mathbb{E}^{b_0,\nu_0}[\langle L^c\rangle_t] \leq K't$ for some other constant $K' > 0$. Likewise,

$$\mathbb{E}^{b_0,\nu_0}\left[\sum_{t:\|\Delta\mathbf{X}_t\|_2\neq 0} \Delta L_t^2\right] = \int_0^t \mathbb{E}^{b_0,\nu_0}\left[\int_{\mathbb{R}_0^d}\left(\frac{d\nu_0}{d\nu}(\mathbf{X}_{s-}, \mathbf{z}) - 1\right)^2 \nu(\mathbf{X}_s, d\mathbf{z})\right] ds$$

is finite due to the aforementioned conditions on ν_0 and ν. Thus L has expected quadratic variation

$$\mathbb{E}^{b_0,\nu_0}[\langle L\rangle_t] = \mathbb{E}^{b_0,\nu_0}\left[\sum_{t:\|\Delta\mathbf{X}_t\|_2\neq 0} \Delta L_t^2 \nu(\mathbf{X}_s, d\mathbf{z})ds + \langle L^c\rangle_t\right] < \infty$$

for any $t > 0$, and is a true $\mathbb{P}^{b_0,\nu_0}$-martingale by Corollary 3 on page 73 of [Protter, 2005]. Then, the Radon-Nikodym term in (9) can be written as

$$\begin{aligned}
&\mathbb{E}^{b_0,\nu_0}\left[\log\left(\frac{d\mathbb{P}_{\mathbf{x}}^{b_0,\nu_0}}{d\mathbb{P}_{\mathbf{x}}^{b,\nu}}((\mathbf{X}_t)_{t\in[0,\delta]})\right)\right] = \mathbb{E}^{b_0,\nu_0}[\log(\mathcal{E}(L_t))] \\
&= \mathbb{E}^{b_0,\nu_0}\left[L_\delta - L_0 - \frac{1}{2}\langle L^c\rangle_\delta + \sum_{t:\Delta\mathbf{X}_t\neq 0}\{\log(1 + \Delta L_t) - \Delta L_t\}\right] \\
&= \mathbb{E}^{b_0,\nu_0}\left[\frac{-1}{2}\int_0^\delta \left\|b_0(\mathbf{X}_t) - b(\mathbf{X}_t) - \int_{\mathbb{R}_0^d}\left(\frac{d\nu_0}{d\nu}(\mathbf{X}_{t-}, \mathbf{z}) - 1\right)\mathbb{1}_{(0,1]}(\|\mathbf{z}\|_2)\mathbf{z}\nu(\mathbf{X}_{t-}, d\mathbf{z})\right\|_2^2 dt\right. \\
&\qquad \left. + \sum_{0\leq t\leq\delta:\Delta\mathbf{X}_t\neq 0}\left\{\log\left(\frac{d\nu_0}{d\nu}(\mathbf{X}_{t-}, \Delta\mathbf{X}_t)\right) - \left(\frac{d\nu_0}{d\nu}(\mathbf{X}_{t-}, \Delta\mathbf{X}_t) - 1\right)\right\}\right] \\
&\leq \delta\left[\frac{1}{2}\left(\|b_0 - b\|_{2,\pi^{b_0,\nu_0}} + \left\|\int_{\mathbb{R}_0^d}\left(\frac{d\nu_0}{d\nu}(\cdot, \mathbf{z}) - 1\right)\mathbb{1}_{(0,1]}(\|\mathbf{z}\|_2)\mathbf{z}\nu(\cdot, d\mathbf{z})\right\|_{2,\pi^{b_0,\nu_0}}\right)^2\right. \\
&\qquad \left. + \left\|\int_{\mathbb{R}_0^d}\left\{\log\left(\frac{d\nu_0}{d\nu}(\cdot, \mathbf{z})\right) - \frac{d\nu_0}{d\nu}(\cdot, \mathbf{z}) + 1\right\}\nu_0(\cdot, d\mathbf{z})\right\|_{1,\pi^{b_0,\nu_0}}\right],
\end{aligned} \tag{12}$$

where the first equality follows from Theorem 2.4 of [Cheridito et al., 2005], the second by definition of $\mathcal{E}$ for jump diffusion processes, and the remainder of the calculation by stationarity and because ν_0 is the compensator of the Poisson random measure driving the jumps of $\mathbf{X}$ under $\mathbb{P}^{b_0,\nu_0}$. The result now follows from (9) and (12). □

Lemma 3. *For each $\delta > 0$ and $f \in C_b(\Omega)$, the collection $\{P_\delta^{b,\nu} f : (b, \nu) \in \Theta\}$ is locally uniformly equicontinuous: for any compact $K \in \Omega$ and $\varepsilon > 0$ there exists $\gamma := \gamma(\varepsilon, K, f, \delta) > 0$ such that*

$$\sup_{(b,\nu)\in\Theta} \sup_{\substack{\mathbf{x},\mathbf{y}\in K: \\ \|\mathbf{x}-\mathbf{y}\|_2<\gamma}} |P_\delta^{b,\nu} f(\mathbf{x}) - P_\delta^{b,\nu} f(\mathbf{y})| < \varepsilon.$$

Proof. Theorem 2.3 of [Wang, 2010] establishes Lipschitz continuity for jump diffusions satisfying (3) using a coupling argument for $f \in \mathcal{B}_b(\Omega)$, the set of bounded, measurable functions. We begin by showing that the conditions of Wang [2010] are satisfied.

In our notation and setting, the condition of Theorem 2.3 of [Wang, 2010] is that for some constant $\beta \in (0, 1)$ there exists a constant $C_\beta > 0$ such that

$$\frac{(1 + \|\mathbf{x} - \mathbf{y}\|_2) \langle b(\mathbf{x}) - b(\mathbf{y}), \mathbf{x} - \mathbf{y}\rangle}{\|\mathbf{x} - \mathbf{y}\|_2} + \frac{(1 + \|\mathbf{x} - \mathbf{y}\|_2) \int_{\|\mathbf{z}\|_2 \leq 1} \|c(\mathbf{x}, \mathbf{z}) - c(\mathbf{y}, \mathbf{z})\|_2^2 M(d\mathbf{z})}{2\|\mathbf{x} - \mathbf{y}\|_2}$$
$$+ (1 + \|\mathbf{x} - \mathbf{y}\|_2) \int_{\|\mathbf{z}\|_2 > 1} \|c(\mathbf{x}, \mathbf{z}) - c(\mathbf{y}, \mathbf{z})\|_2 M(d\mathbf{z}) + (1 + \|\mathbf{x} - \mathbf{y}\|_2) C_\beta \leq 2$$

whenever $\|\mathbf{x} - \mathbf{y}\|_2 < \beta$ and where $\langle \cdot, \cdot \rangle$ denotes the usual Euclidean inner product. By (3), the first two summands on the LHS can be bounded by $\beta(1+\beta)\sqrt{C_1}$ and $\beta(1+\beta)C_1/2$, respectively. The fourth is trivially bounded by $(1+\beta)C_\beta$. By Jensen's inequality, (3) and (7), the third term can be bounded by $\beta(1 + \beta)\sqrt{C_1 C_5}$. Hence, the whole LHS can be bounded by

$$(1 + \beta) \left[\beta\sqrt{C_1}\left(1 + \sqrt{C_5}\right) + \frac{\beta C_1}{2} + C_\beta\right]$$

which can clearly be made arbitrarily small by choosing both β and C_β to be sufficiently small. This choice can be made uniformly due to the uniform bounds on the Lipschitz constant C_1 and the total mass constraint C_5.

Now, the Lipschitz constant in [Wang, 2010, Theorem 2.3] is of the form

$$2(1 + \beta) \left(\frac{1}{C_\beta \Delta} + \frac{1}{\beta}\right) \|f\|.$$

Since β and C_β can be chosen uniformly in Θ, and Δ and $\|f\|$ are constants, uniform equicontinuity holds. ☐

The remainder of the proof follows as in [van der Meulen and van Zanten, 2013]. It suffices to show that for $f \in C_b(\Omega)$ and $B := \{(b, \nu) \in \Theta : \|P_\delta^{b,\nu} f - P_\delta^{b_0,\nu_0} f\|_{1,\rho} > \varepsilon\}$ we have $\Pi(B|\mathbf{x}_{0:n}) \to 0$ with $\mathbb{P}^{b_0,\nu_0}$-probability 1. To that end we fix $f \in \mathrm{Lip}(\Omega)$ and $\varepsilon > 0$ and thus the set B. Lemma 2 implies that Lemma 5.2 of [van der Meulen and van Zanten, 2013] holds, so that if, for measurable subsets $C_n \subset \Theta$, there exists $c > 0$ such that

$$e^{nc} \int_{C_n} \pi^{b,\nu}(\mathbf{x}_0) \prod_{i=1}^{n} p_\delta^{b,\nu}(\mathbf{x}_{i-1}, \mathbf{x}_i) \Pi(db, d\nu) \to 0$$

$\mathbb{P}^{b_0,\nu_0}$-a.s. then $\Pi(C_n|\mathbf{x}_{0:n}) \to 0$ $\mathbb{P}^{b_0,\nu_0}$-a.s. as well. Likewise, Lemma 3 implies Lemma 5.3 of [van der Meulen and van Zanten, 2013]: there exists a compact subset $K \subset \Omega$, $N \in \mathbb{N}$ and compact, connected sets $I_1, \ldots, I_N$ that cover K such that

$$B \subset \bigcup_{j=1}^{N} B_j^+ \cup \bigcup_{j=1}^{N} B_j^-,$$

where

$$B_j^+ := \left\{(b, \nu) \in \Theta : P_\delta^{b,\nu} f(\mathbf{x}) - P_\delta^{b_0,\nu_0} f(\mathbf{x}) > \frac{\varepsilon}{4\nu(K)} \text{ for every } \mathbf{x} \in I_j\right\},$$
$$B_j^- := \left\{(b, \nu) \in \Theta : P_\delta^{b,\nu} f(\mathbf{x}) - P_\delta^{b_0,\nu_0} f(\mathbf{x}) < \frac{-\varepsilon}{4\nu(K)} \text{ for every } \mathbf{x} \in I_j\right\}.$$

Thus it is only necessary to show $\Pi(B_j^{\pm}|\mathbf{x}_{0:n}) \to 0$ $\mathbb{P}^{b_0,\nu_0}$-almost surely. Define the stochastic process

$$D_n := \left(\int_{B_j^+} \pi^{b,\nu}(\mathbf{x}_0) \prod_{i=1}^{n} p_\delta^{b,\nu}(\mathbf{x}_{i-1}, \mathbf{x}_i) \Pi(db, d\nu) \right)^{1/2}.$$

Now $D_n \to 0$ exponentially fast as $n \to \infty$ by an argument identical to that used to prove Theorem 3.5 of [van der Meulen and van Zanten, 2013]. The same is also true of the analogous stochastic process defined by integrating over B_j^-, which completes the proof. ☐

4 An example prior

The conditions of Theorem 1 are verifiable in the sense that they do not depend on intractable quantities (with the exception of Assumption 1), but it is not immediately clear whether a prior Π satisfying its assumptions exists, in particular in the infinite dimensional setting. The following example demonstrates that there is at least one family of priors which satisfies these assumptions: independent discrete net priors of Ghosal et al. [1997] for $b(\cdot)$ and $c(\cdot, \cdot)$, and a further, independent Dirichlet process mixture model prior [Lo, 1984] for $M(\cdot)$. Discrete net priors were also used in both van der Meulen and van Zanten [2013] and Gugushvili and Spreij [2014] to demonstrate the existence of priors for nonparametric inference of drifts for diffusions.

Firstly, let Θ_b be a collection of uniformly Lipschitz functions from Ω to $\mathbb{R}^d$, each satisfying (5) for some (not necessarily uniform) constants C_3 and C_4. Let $\Theta_b^{(m)} := \{b|_{\overline{B_0(m)}} : b \in \Theta_b\}$ be the set of restriction in Θ_b to the closed ball of radius m centred at the origin. By uniform equicontinuity and the Arzelà-Ascoli theorem, $\Theta_b^{(m)}$ is totally bounded in the uniform norm. Hence, for every n, it is possible to construct a finite ε_n-net $\Theta_b^{(m,n)}$ over $\Theta_b^{(m)}$, where $\{\varepsilon_n\}_{n\in\mathbb{N}}$ is a sequence of strictly positive numbers tending to 0. In other words, $\Theta_b^{(m,n)}$ is a finite set with the property that every element of $\Theta_b^{(m)}$ is within distance ε_n of some element of $\Theta_b^{(m,n)}$ in the supremum norm. Finally, every $b \in \Theta_b^{(m,n)}$ is extended to Ω by setting $b(\mathbf{x}) = b(P_{\overline{B_0(m)}}\mathbf{x}) - \mathbf{x} + P_{\overline{B_0(m)}}\mathbf{x}$ outside $\overline{B_0(m)}$, where $P_{\overline{B_0(m)}}$ is the orthogonal projection onto $\overline{B_0(m)}$. Now, a discrete net prior is constructed by fixing two probability mass functions on $\mathbb{N}$, $\{p_m\}_{m\in\mathbb{N}}$ and $\{q_n\}_{n\in\mathbb{N}}$, both of which assign positive mass to every positive integer. Then, a draw from the prior is generated by sampling $m \sim p_m$ and $n \sim q_n$, followed by $b|m, n \sim U(\Theta_b^{(m,n)})$. Samples from this prior are bounded, uniformly Lipschitz continuous, and satisfy (5) by construction.

Now let $J \subset \mathbb{R}^d$ be a fixed, compact domain including the origin, and let Θ_c be a set of uniformly Lipschitz continuous functions $c : \Omega \times J \mapsto J$ which satisfy the following:

1. $c(\cdot, 0) \equiv 0$,
2. $c(\mathbf{x}, \cdot) : J \mapsto J$ is a surjection for each $\mathbf{x} \in \Omega$,
3. for any $c \in \Theta_c$, any $\mathbf{x} \in \Omega$ and $\mathbf{z} \in \mathbb{R}_0^d$, there exists an open ball of strictly positive radius centred at $\mathbf{z}$, $B_{\mathbf{z}}(\varepsilon)$, such that $c(\mathbf{x}, \mathbf{z}) \neq c(\mathbf{x}, \boldsymbol{\xi})$ for any $\mathbf{z} \neq \boldsymbol{\xi} \in B_{\mathbf{z}}(\varepsilon)$,
4. for each $c \in \Theta_c$ there exists $K_c > 0$ such that

$$\sup_{\mathbf{x}\in\Omega} \left\{ \sup_{\mathbf{z}\in J} \{c(\mathbf{x}, \mathbf{z})\} \right\} = \sup_{\mathbf{x}:\|\mathbf{x}\|_2 \leq K_c} \left\{ \sup_{\mathbf{z}\in J} \{c(\mathbf{x}, \mathbf{z})\} \right\}, \tag{13}$$

   and likewise for infima.

Condition 2. guarantees that $\nu(\mathbf{x}, d\mathbf{z})$ has a positive density everywhere whenever $M(d\mathbf{z})$ does, while condition 3. rules out atoms in $\nu(\mathbf{x}, d\mathbf{z})$. Let $\Theta_c^{(m)} := \{c|_{\overline{B_0(m)}} : c \in \Theta_c\}$ be the set of

restrictions of the first coordinate to the ball $\overline{B_0(m)} \subset \Omega$, and let $\Theta_c^{(m,n)}$ be a $\tilde{\varepsilon}_n$-net over $\Theta_c^{(m)}$ for a strictly positive sequence $\tilde{\varepsilon}_n \searrow 0$. Each element of $\Theta_c^{(m,n)}$ can again be extended to a function on the whole $\Omega \times J$ by setting $c(\mathbf{x}, \mathbf{z}) = c(P_{\overline{B_0(m)}}\mathbf{x}, \mathbf{z})$ outside $\overline{B_0(m)}$, where $P_{\overline{B_0(m)}}$ denotes the orthogonal projection to $\overline{B_0(m)}$ as before. An independent discrete net can be used to define a prior for $c(\cdot, \cdot)$, by specifying two probability mass functions $\{\tilde{p}_m\}_{m\in\mathbb{N}}$ and $\{\tilde{q}_n\}_{n\in\mathbb{N}}$, both assigning positive mass to all positive integers, and sampling draws analogously to the discrete net prior on Θ_b.

Finally take the prior for the intensity measure $M(\cdot)$ to be a Dirichlet process mixture model [Lo, 1984]. Let $\phi_\tau(\mathbf{z})$ denote the d-dimensional centred Gaussian density with covariance matrix $\tau^{-1}\mathbb{I}_{d\times d}$ restricted to J, and renormalised to be a probability density. Let F be a probability measure on $(0, \infty)$ assigning positive mass to all non-empty open sets, and let $\mathrm{DP}(\zeta)$ denote the law of a Dirichlet process [Ferguson, 1973] with the mean measure $\zeta \in \mathcal{M}_f(J)$, which is taken to be a probability measure with a finite first moment, independent of F. Let $\mathcal{D}_\Upsilon(J)$ denote the space of continuous, positive densities on J with total mass at most $\Upsilon > 0$. The Dirichlet process mixture model on $\mathcal{D}_\Upsilon(J)$ with truncated Gaussian mixture kernel ϕ_τ and mixing distribution $U(0, \Upsilon) \otimes F \otimes \mathrm{DP}(\zeta)$ is specified via the following sampling procedure:

1. Sample $P \sim \mathrm{DP}(\zeta)$. Then P is a discrete probability measure on $\mathbb{R}^d$ with countably many atoms with $\mathrm{DP}(\zeta)$-probability 1 [Ferguson, 1973]. Let $\mathbf{z}_1, \mathbf{z}_2, \ldots$ denote these atoms in some fixed ordering.
2. Sample IID copies $\tau_1, \tau_2, \ldots \sim F$.
3. Sample $\alpha \sim U(0, \Upsilon)$.
4. Set $M(d\mathbf{z}) = \alpha \sum_{j=1}^{\infty} P(\mathbf{z}_j)\phi_{\tau_j}(\mathbf{z} - \mathbf{z}_j)d\mathbf{z}$.

Note that samples are finite measures with strictly positive densities on J, which also means they have second moments because J is compact.

Sampling all three components, $b(\cdot)$, $c(\cdot, \cdot)$ and $M(\cdot)$ independently from the priors specified above yields draws which almost surely satisfy (3) by uniform Lipschitz continuity of b and c, as well as a uniform bound on the total mass of M:

$$\begin{aligned}\|b(\mathbf{x}) - b(\mathbf{y})\|_2^2 &+ \int_J \|c(\mathbf{x}, \mathbf{z}) - c(\mathbf{y}, \mathbf{z})\|_2^2 M(d\mathbf{z}) \\ &\leq C_b\|\mathbf{x} - \mathbf{y}\|_2^2 + \int_J C_c\|\mathbf{x} - \mathbf{y}\|_2^2 M(d\mathbf{z}) \leq (C_b + \Upsilon C_c)\|\mathbf{x} - \mathbf{y}\|_2^2.\end{aligned}$$

Condition (4) is immediate from the uniform Lipschitz continuity of c, and (5) holds by construction of the prior for b. The requirement that $c(\cdot, 0) \equiv 0$ holds by construction. Finally, $\nu(\mathbf{x}, d\mathbf{z}) = M(c^*(\mathbf{x}, d\mathbf{z}))$ is a finite measure for each $\mathbf{x}$ because M is finite and $c^*(\mathbf{x}, \mathbf{z})$ is a finite union of points by non-constancy, Lipschitz continuity and compactness of J. The Radon-Nikodym derivative $\frac{d\nu}{d\nu_0}$ exists for the same reason, and is bounded both from above and away from 0 by compactness of J and (13). Thus, the conditions of Definition 1 are fulfilled.

It remains to verify that (8) holds for this product prior. This will be achieved by controlling the three π^{b_0,ν_0}-norms separately, and showing that samples which result in all three taking arbitrarily small values are drawn with positive probability.

First, fix $b_0 \in \Theta_b$, $c_0 \in \Theta_c$ and $M_0 \in \mathcal{D}_\Upsilon(J)$, as well as $\varepsilon > 0$, and define

$$\|b\|_{m,\infty} := \sup_{\|\mathbf{x}\|_2 \leq m} \|b(\mathbf{x})\|_\infty. \tag{14}$$

Note that $\|\cdot\|_{m,\infty}$ is well defined for Lipschitz functions because they are locally bounded. Then

$$\begin{aligned}\|b_0 - b\|^2_{2,\pi^{b_0,\nu_0}} &\leq \|b_0 - b\|^2_{m,\infty} + \int_{\|\mathbf{x}\|_2 > m} \|b_0(\mathbf{x}) - b(\mathbf{x})\|_2^2 \pi^{b_0,\nu_0}(\mathbf{x}) d\mathbf{x} \\ &\leq \|b_0 - b\|^2_{m,\infty} + \int_{\|\mathbf{x}\|_2 > m} \|b_0(\mathbf{0}) + b(\mathbf{0}) + 2C_1 \mathbf{x}\|_2^2 \pi^{b_0,\nu_0}(\mathbf{x}) d\mathbf{x},\end{aligned}$$

by Lipschitz continuity. Now, choose m to be large enough that the second term on the RHS is bounded above by $\varepsilon/8$, which can be done because π^{b_0,ν_0} has second moments by Proposition 1. Likewise, the first term can be bounded by $\varepsilon/8$ by choosing n large enough that $\varepsilon_n \leq \varepsilon/8$. Note that by construction, the probability of sampling a corresponding function b from the prior is at least $p_m q_n > 0$, and that for such a b we have

$$\|b_0 - b\|_{2,\pi^{b_0,\nu_0}} \leq \sqrt{2\varepsilon/8} = \sqrt{\varepsilon}/2.$$

For the second norm, an elementary calculation using Jensen's inequality and the fact that $\|\mathbf{z}\|_2 \leq 1$ yields that

$$\begin{aligned}&\left\| \int_J \left[\frac{d\nu_0}{d\nu}(\cdot, \mathbf{z}) - 1 \right] \mathbb{1}_{(0,1]}(\|\mathbf{z}\|_2) \mathbf{z} \nu(\cdot, d\mathbf{z}) \right\|_{2,\pi^{b_0,\nu_0}} \\ &\leq \int_\Omega \nu(\mathbf{x}, B_0(1)) \int_J \left(\frac{\nu_0(\mathbf{x}, \mathbf{z})^2 - 2\nu_0(\mathbf{x}, \mathbf{z})\nu(\mathbf{x}, \mathbf{z}) + \nu(\mathbf{x}, \mathbf{z})^2}{\nu(\mathbf{x}, \mathbf{z})^2} \right) \nu(\mathbf{x}, d\mathbf{z}) \pi^{b_0,\nu_0}(\mathbf{x}) d\mathbf{x} \\ &\leq \int_\Omega \pi^{b_0,\nu_0}(\mathbf{x}) \nu(\mathbf{x}, B_0(1)) \Bigg(\left| \int_{\mathbf{z}:\|\mathbf{z}\|_2 \leq 1} \frac{d\nu_0}{d\nu}(\mathbf{x}, \mathbf{z}) \nu_0(\mathbf{x}, d\mathbf{z}) - \nu_0(\mathbf{x}, B_0(1)) \right| \\ &\qquad\qquad + |\nu(\mathbf{x}, B_0(1)) - \nu_0(\mathbf{x}, B_0(1))| \Bigg) d\mathbf{x} \\ &= \int_{\mathbf{x}:\|\mathbf{x}\|_2 \leq m} \pi^{b_0,\nu_0}(\mathbf{x}) \nu(\mathbf{x}, B_0(1)) \Bigg(\left| \int_{\mathbf{z}:\|\mathbf{z}\|_2 \leq 1} \frac{d\nu_0}{d\nu}(\mathbf{x}, \mathbf{z}) \nu_0(\mathbf{x}, d\mathbf{z}) - \nu_0(\mathbf{x}, B_0(1)) \right| \\ &\qquad\qquad + |\nu(\mathbf{x}, B_0(1)) - \nu_0(\mathbf{x}, B_0(1))| \Bigg) d\mathbf{x} \\ &+ \int_{\mathbf{x}:\|\mathbf{x}\|_2 > m} \pi^{b_0,\nu_0}(\mathbf{x}) \nu(\mathbf{x}, B_0(1)) \Bigg(\left| \int_{\mathbf{z}:\|\mathbf{z}\|_2 \leq 1} \frac{d\nu_0}{d\nu}(\mathbf{x}, \mathbf{z}) \nu_0(\mathbf{x}, d\mathbf{z}) - \nu_0(\mathbf{x}, B_0(1)) \right| \\ &\qquad\qquad + |\nu(\mathbf{x}, B_0(1)) - \nu_0(\mathbf{x}, B_0(1))| \Bigg) d\mathbf{x}. \end{aligned} \tag{15}$$

Both $\nu(\mathbf{x}, \cdot)$ and $\nu_0(\mathbf{x}, \cdot)$ are finite measures with Radon-Nikodym derivative bounded from above and away from 0. Thus, finiteness of π^{b_0,ν_0} ensures that the second integral on the RHS can be made arbitrarily small by choosing large enough m. Now consider

$$\begin{aligned}&|\nu(\mathbf{x}, B_0(1)) - \nu_0(\mathbf{x}, B_0(1))| = |M(c^*(\mathbf{x}, B_0(1))) - M_0(c_0^*(\mathbf{x}, B_0(1)))| \\ &\leq |M(c^*(\mathbf{x}, B_0(1))) - M(c_0^*(\mathbf{x}, B_0(1)))| + |M(c_0^*(\mathbf{x}, B_0(1))) - M_0(c_0^*(\mathbf{x}, B_0(1)))|,\end{aligned}$$

and note that if $\|c - c_0\|_{m,\infty} \leq \gamma_1$ then

$$c^*(\mathbf{x}, B_0(1)) \subseteq c_0^* \left(\mathbf{x}, \left\{ \boldsymbol{\xi} \in J : \inf_{\mathbf{y} \in B_0(1)} \{ \|\boldsymbol{\xi} - \mathbf{y}\|_\infty \} \leq \gamma_1 \right\} \right) \tag{16}$$

for any $\mathbf{x} : \|\mathbf{x}\|_2 \leq m$. The sets on the RHS are decreasing with decreasing $\gamma_1 > 0$ and of finite M-mass, so that continuity of measure gives $|M(c^*(\mathbf{x}, B_0(1))) - M(c_0^*(\mathbf{x}, B_0(1)))| \leq \gamma_2$ for some γ_2, which decreases to 0 as $\gamma_1 \searrow 0$. Likewise,

$$|M(c_0^*(\mathbf{x}, B_0(1))) - M_0(c_0^*(\mathbf{x}, B_0(1)))| \leq \|M - M_0\|_\infty.$$

Hence

$$|\nu(\mathbf{x}, B_0(1)) - \nu_0(\mathbf{x}, B_0(1))| \leq \gamma_2 + \|M - M_0\|_\infty,$$

which can be made arbitrarily small by first choosing a sufficiently large m, then a sufficiently small γ_2 as well as $c : \|c - c_0\|_{m,\infty} < \gamma_2$, and finally an $M : \|M - M_0\|_\infty < \gamma_3$ for sufficiently small γ_3.

Similarly, (16) gives that

$$\frac{d\nu_0}{d\nu}(\mathbf{x}, \mathbf{z}) = \frac{M_0(c_0^*(\mathbf{x}, \mathbf{z}))}{M(c^*(\mathbf{x}, \mathbf{z}))} \leq \frac{M_0(c^*(\mathbf{x}, \{\boldsymbol{\xi} \in J : \|\boldsymbol{\xi} - \mathbf{z}\|_\infty \leq \gamma_1\}))}{M(c^*(\mathbf{x}, \mathbf{z}))}$$

for $\mathbf{x} : \|\mathbf{x}\|_2 \leq m$ whenever $\|c - c_0\|_{m,\infty} < \gamma_1$. Hence taking such a c, as well as $M : \|M - M_0\|_\infty < \gamma_3$, and using continuity of measure yields the estimate

$$\frac{d\nu_0}{d\nu}(\mathbf{x}, \mathbf{z}) \leq \frac{M(c^*(\mathbf{x}, \mathbf{z})) + \gamma_3 + \gamma_4}{M(c^*(\mathbf{x}, \mathbf{z}))}$$

for some $\gamma_4 \searrow 0$ as $\gamma_1 \searrow 0$. The denominator is bounded from below, so that

$$\frac{d\nu_0}{d\nu}(\mathbf{x}, \mathbf{z}) \leq 1 + \frac{\gamma_3 + \gamma_4}{\inf_{\|\mathbf{x}\|_2 \leq m, \|\mathbf{z}\|_2 \leq 1}\{\nu(\mathbf{x}, \mathbf{z})\}},$$

which can be made arbitrarily close to 1 by choosing small enough γ_3 and γ_4. An analogous lower bound follows by reversing the roles of ν and ν_0, and lower bounding the Radon-Nikodym derivative instead. Thus

$$(1 - \gamma_3 - \gamma_4)\nu_0(\mathbf{x}, B_0(1)) \leq \int_{\mathbf{z}:\|\mathbf{z}\|_2 \leq 1} \frac{d\nu_0}{d\nu}(\mathbf{x}, \mathbf{z})\nu_0(\mathbf{x}, d\mathbf{z}) \leq \nu_0(\mathbf{x}, B_0(1))(1 + \gamma_3 + \gamma_4),$$

so that

$$\Big| \int_{\mathbf{z}:\|\mathbf{z}\|_2 \leq 1} \frac{d\nu_0}{d\nu}(\mathbf{x}, \mathbf{z})\nu_0(\mathbf{x}, d\mathbf{z}) - \nu_0(\mathbf{x}, B_0(1)) \Big| \leq \gamma_3 + \gamma_4.$$

Taken together, the above bounds imply that (15) can be bounded by $\sqrt{\varepsilon}/2$ by first choosing a large enough m, and then c and M such that $\|c - c_0\|_{m,\infty} < \gamma$ and $\|M - M_0\|_{m,\infty} < \gamma$ for sufficiently small $\gamma > 0$. Fix n such that $\tilde{\varepsilon}_n \leq \gamma$. Then a suitable c is sampled from the prior with probability at least $\tilde{p}_m \tilde{q}_n > 0$. The probability of sampling a suitable M is also positive, because by Theorem 1 of Bhattacharya and Dunson [2012], the support of the Dirichlet process mixture model is dense in $\mathcal{D}_\Upsilon(J)$.

The third norm in (8) can be treated identically to the second, because $x \mapsto \log(x)$ is continuous and J is compact. Hence, its value is also bounded by $\varepsilon/2$ with strictly positive Π-probability. Thus, with positive Π-probability

$$\frac{1}{2}\left(\|b_0 - b\|_{2,\pi^{b_0,\nu_0}} + \Big\| \int_{\mathbb{R}_0^d} \left[\frac{d\nu_0}{d\nu}(\cdot, \mathbf{z}) - 1 \right] \mathbb{1}_{(0,1]}(\|\mathbf{z}\|_2)\mathbf{z}\nu(\cdot, d\mathbf{z}) \Big\|_{2,\pi^{b_0,\nu_0}} \right)^2$$
$$+ \Big\| \int_{\mathbb{R}_0^d} \left[\log\left(\frac{d\nu_0}{d\nu}(\cdot, \mathbf{z}) \right) - \frac{d\nu_0}{d\nu}(\cdot, \mathbf{z}) + 1 \right] \nu_0(\cdot, d\mathbf{z}) \Big\|_{1,\pi^{b_0,\nu_0}} < \frac{1}{2}\left(\frac{\sqrt{\varepsilon}}{2} + \frac{\sqrt{\varepsilon}}{2} \right)^2 + \frac{\varepsilon}{2} = \varepsilon,$$

and hence (8) holds.

5 Discussion

In this paper we have shown that posterior consistency for identifiable, nonparametric Bayesian inference of drift and jump coefficients of jump diffusions from discrete data holds under criteria which can be readily checked in practice. Identifiability itself seems difficult to verify beyond the scope of gradient-type diffusions, for which it was established along with posterior consistency in [Gugushvili and Spreij, 2014]. Products of discrete net priors and Dirichlet process mixture models were shown to satisfy the conditions for consistency, provided that identifiability holds. It seems likely that in a case where identifiability fails but all other consistency criteria hold, the posterior will converge to be supported on the subset of pairs (b, ν) that give rise to the semigroup generating the data, with weights proportional to the prior densities of the pairs, at least subject to the set of these pairs being sufficiently regular. However, this conjecture has not been verified rigorously.

Our results shares the limitation of [van der Meulen and van Zanten, 2013, Gugushvili and Spreij, 2014] of requiring priors to assign full mass to sets of functions for which the Lipschitz condition (3) holds uniformly. This rules out many widely used families of priors, but counterexamples exist to show that without it, uniform equicontinuity fails even for one dimensional unit diffusions [Matthias Birkner, personal communication]. It seems clear that an entirely different approach is needed, if consistency results are to be established without a uniform equicontinuity condition.

A further limitation of [van der Meulen and van Zanten, 2013, Gugushvili and Spreij, 2014] is that of being established for a weak topology, for which the martingale approach of [Walker, 2004, Lijoi et al., 2004] is well suited. A testing approach, such as that of [Ghosal and van der Vaart, 2007], would yield convergence in a stronger topology as well as rates of convergence, but it is not clear how to adapt their results to the diffusion or jump diffusion settings. Currently, results in this direction are only available for continuously observed scalar diffusions [van der Meulen et al., 2006, Panzar and van Zanten, 2009, Pokern et al., 2013], as well as discretely observed scalar diffusions on a compact interval [Nickl and Söhl, 2017]. However, these rely, respectively, on the continuity of the observation and on a tractable representation of the stationary density, neither of which is available more generally.

Practical implementation of inference algorithms is beyond the scope of this paper, but we note that algorithms based on exact simulation for jump diffusions are available, at least in the scalar case [Casella and Roberts, 2011, Gonçalves, 2011, Pollock et al., 2017]. Exact simulation of jump diffusions is an active area of research [Gonçalves and Roberts, 2013, Pollock, 2015, Pollock et al., 2016] and well suited for applications in Monte Carlo inference algorithms, with preliminary results in the continuous diffusion setting indicating that nonparametric algorithms can be feasibly implemented [Papaspiliopoulos et al., 2012, van Zanten, 2013, van der Meulen et al., 2014]. As a final remark, we note that presently such algorithms are only available for processes with jumps driven by compound Poisson processes of finite intensity, and with coefficients satisfying regularity assumptions comparable to those in Proposition 1. Thus our Theorem 1 brings the theory on nonparametric posterior consistency in line with current state of the art algorithms in one dimension, and anticipates development of comparable methods in higher dimensions.

Acknowledgements

The authors are grateful to Matthias Birkner for comments leading to substantial improvements in this manuscript. Jere Koskela was supported by EPSRC as part of the MASDOC DTC at the University of Warwick. Grant No. EP/HO23364/1. Paul Jenkins is supported in part by EPSRC grant EP/L018497/1.

References

K. K. Aase and P. Guttorp. Estimation in models for security prices. *Scand. Actuarial J.*, pages 211–224, 1987.

Y. Aït-Sahalia. Closed-form likelihood expansions for multivariate diffusions. *Ann. Stat.*, 36(2): 906–937, 2008.

D. Applebaum. *Lévy processes and stochastic calculus.* Cambridge studies in advanced mathematics. Cambridge University Press, 2004.

S. P. Au, A. H. Haddad, and V. H. Poor. A state estimation algorithm for linear systems driven simultaneously by Weiner and Poisson processes. *IEEE Trans. Aut. Control*, Ac-27(3): 617–626, 1982.

I. Bardhan and X. Chao. Pricing options on securities with discontinuous returns. *Stoch. Proc. Appl.*, 48(1):123–137, 1993.

J. Bertoin and J.-F. Le Gall. Stochastic flows associated to coalescent processes. *Probab. Theory Related Fields*, 126:261–288, 2003.

A. Bhattacharya and D. B. Dunson. Strong consistency of nonparametric Bayes density estimation on compact metric spaces with applications to specific manifolds. *Ann. Inst. Stat. Math.*, 64:687–714, 2012.

M. Birkner, J. Blath, M. Möhle, M. Steinrücken, and J. Tams. A modified lookdown construction for the Ξ-Fleming-Viot process with mutation and populations with recurrent bottlenecks. *Alea*, 6:25–61, 2009.

M. Bladt and Sørensen. Statistical inference for discretely observed Markov jump processes. *J R Stat. Soc. Ser. B*, 67(3):395–410, 2005.

B. A. Bodo, M. E. Thompson, and T. E. Unny. A review of stochastic differential equations for applications in hydrology. *Stoch. Hydrol. Hydraul.*, 2:81–100, 1987.

B. Casella and G. O. Roberts. Exact simulation of jump-diffusion processes with Monte Carlo applications. *Methodol. Comput. Appl. Probab.*, 13:449–473, 2011.

L. Chen and D. Filipović. A simple model for credit migration and spread curves. *Finance Stochast.*, 9:211–231, 2005.

P. Cheridito, D. Filipović, and M. Yor. Equivalent and absolutely continuous measure changes for jump-diffusion processes. *Ann. Appl. Probab.*, 15(3):1713–1732, 2005.

F. Comte, V. Genon-Catalot, and Y. Rozenholc. Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13:514–543, 2007.

A. Dalalyan and M. Reiß. Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probab. Theory Related Fields*, 137:25–47, 2007.

P. Diaconis and D. Freedman. On the consistency of Bayes estimates. with a discussion and a rejoinder by the authors. *Ann. Statist.*, 14:1–67, 1986.

R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge studies in advanced mathematics.* Cambridge University Press, revised reprint of the 1989 original edition, 2002.

T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *Ann. Stat.*, 1(2):209–230, 1973.

D. Filipović, P. Cheridito, and R. L. Kimmel. Market price of risk specifications for affine models: theory and evidence. *J. Financ. Econ.*, 83(1):123–170, 2007.

S. Fornaro. *Regularity properties for second order partial differential operators with unbounded coefficients*. PhD thesis, Università del Salento, 2004.

S. Ghosal and Y. Tang. Bayesian consistency for Markov processes. *Sankhya*, 68:227–239, 2006.

S. Ghosal and A. W. van der Vaart. Convergence rates of posterior distributions for non-i.i.d observations. *Ann. Statist.*, 35:192–223, 2007.

S. Ghosal, J. K. Ghosh, and R. V. Ramamoorthi. Non-informative priors via sieves and packing numbers. In S. Panchapakesan and N. Balakrishnan, editors, *Advances in statistical decision theory and applications*. Birkhäuser, 1997.

E. Gobet, M. Hoffmann, and M. Reiß. Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.*, 32:2223–2253, 2004.

F. B. Gonçalves. *Exact simulation and Monte Carlo inference for jump-diffusion processes*. PhD thesis, University of Warwick, 2011.

F. B. Gonçalves and G. O. Roberts. Exact simulation problems for jump-diffusions. *Methodol. Comput. Appl. Probab.*, 16(4):907–930, 2013.

S. Gugushvili and P. Spreij. Non-parametric Bayesian drift estimation for stochastic differential equations. *Lith. Math. J.*, 54(2):127–141, 2014.

J. Jacod. Non-parametric kernel estimation of the coefficients of a diffusion. *Scand. J. Statist.*, 27:83–96, 2000.

G. Kallianpur. Differential-equations models for spatially distributed neurons and propagation of chaos for interacting systems. *Math. Biosc.*, 112:207–224, 1992.

G. Kallianpur and J. Xiong. Asymptotic behaviour of a system of interacting nuclear-space-valued stochastic differential equations driven by Poisson random measures. *Appl. Math. Opt.*, 30:175–201, 1994.

V. N. Kolokoltsov. On Markov processes with decomposable pseudo-differential generators. *Stoch. Stoch. Rep.*, 76(1):1–44, 2004.

A. Lijoi, I. Prünster, and S. G. Walker. Extending Doob's consistency theorem to nonparametric densities. *Bernoulli*, 10(4):651–663, 2004.

A. Y. Lo. On a class of Bayesian nonparametric estimates. 1. density estimates. *Ann. Statist.*, 12:351–357, 1984.

H. Masuda. Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stoch. Proc. Appl.*, 117:35–56, 2007.

H. Masuda. Erratum to "Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps". *Stoch. Proc. Appl.*, 119:676–678, 2009.

R. C. Merton. Option pricing when underlying stock returns are discontinuous. *J. Financ. Econ.*, 3:125–144, 1976.

R. Nickl and J. Söhl. Nonparametric Bayesian contraction rates for discretely observed scalar diffusions. *Ann. Stat.*, to appear, 2017.

L. Panzar and H. van Zanten. Nonparametric Bayesian inference for ergodic diffusions. *J. Statist. Plann. Inference*, 139:4193–4199, 2009.

O. Papaspiliopoulos, Y. Pokern, G. O. Roberts, and A. M. Stuart. Nonparametric estimation of diffusions: a differential equations approach. *Biometrika*, 99:511–531, 2012.

Y. Pokern, A. M. Stuart, and H. van Zanten. Posterior consistency via precision operators for nonparametric drift estimation in SDEs. *Stochastic Process. Appl.*, 123:603–628, 2013.

M. Pollock. On the exact simulation of (jump) diffusion bridges. In *Proceedings of the 2015 Winter Simulation Conference*. IEEE Press, 2015.

M. Pollock, A. M. Johansen, and G. O. Roberts. On the exact and ε-strong simulation of (jump) diffusions. *Bernoulli*, 22:794–856, 2016.

M. Pollock, A. M. Johansen, and G. O. Roberts. Particle filtering for partially observed jump diffusions. *In preparation*, 2017.

P. Protter. *Stochastic integration and differential equations*. Stochastic modelling and applied probability. Springer-Verlag, second edition, 2005.

R. L. Schilling and J. Wang. Some theorems on Feller processes: transience, local times and ultracontractivity. *T. Am. Math. Soc.*, 365(6):3255–3286, 2013.

E. Schmisser. Penalized nonparametric drift estimation for multidimensional diffusion processes. *Statistics*, 47:61–84, 2013.

O. Stramer and R. L. Tweedie. Existence and stability of weak solutions to stochastic differential equations with non-smooth coefficients. *Stat. Sinica*, 7:577–593, 1997.

Y. Tang and S. Ghosal. Posterior consistency of Dirichlet mixtures for estimating a transition density. *J. Statist. Plann. Inference*, 137:1711–1726, 2007.

F. van der Meulen and H. van Zanten. Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli*, 19(1):44–63, 2013.

F. van der Meulen, A. W. van der Vaart, and H. van Zanten. Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli*, 12(5):863–888, 2006.

F. van der Meulen, M. Schauer, and H. van Zanten. Reversible jump MCMC for non-parametric drift estimation for diffusion processes. *Comput. Stat. Data An.*, 71:615–632, 2014.

H. van Zanten. Nonparametric Bayesian methods for one-dimensional diffusion models. *Math. Biosci.*, 243(2):215–222, 2013.

S. Walker. New approaches to Bayesian consistency. *Ann. Statist.*, 32:2028–2043, 2004.

J. Wang. Regularity of semigroups generated by Lévy type operators via coupling. *Stoch. Proc. Appl.*, 120(9):1680–1700, 2010.

J. Yu. Closed-form likelihood approximation and estimation of jump-diffusions with an application to the realignment risk of the Chinese Yuan. *J. Econometrics*, 141:1245–1280, 2007.